Supporting Evidence-Informed Policymaking:

A Case Study In Visualizing Twitter Networks

By Bruce Hoppe, Sarah Lucas, and Claire Reinelt

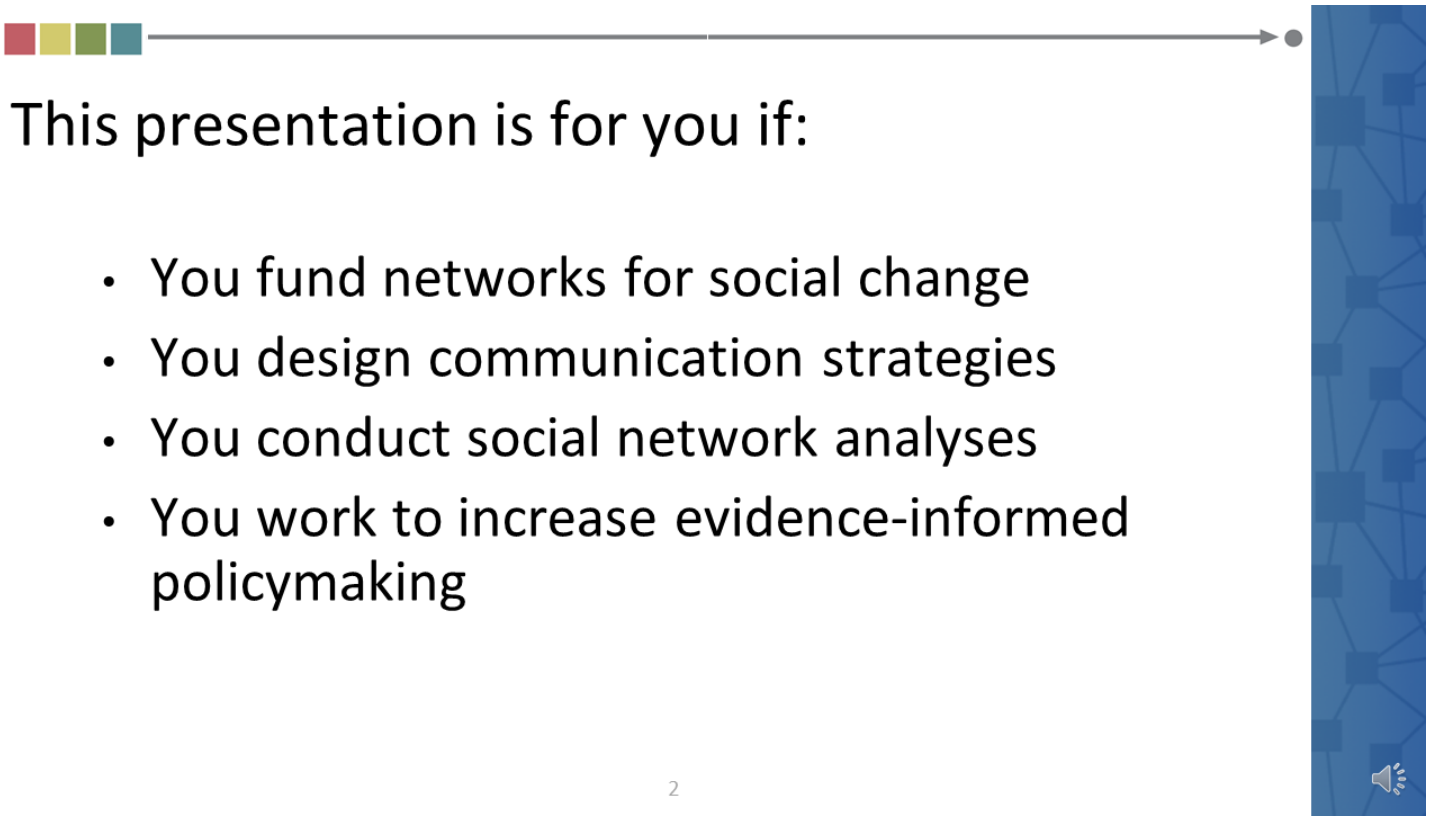
With Support from the William and Flora Hewlett Foundation

March 27, 2019

Supporting Evidence-Informed Policymaking: A Case Study in Twitter Network Visualization

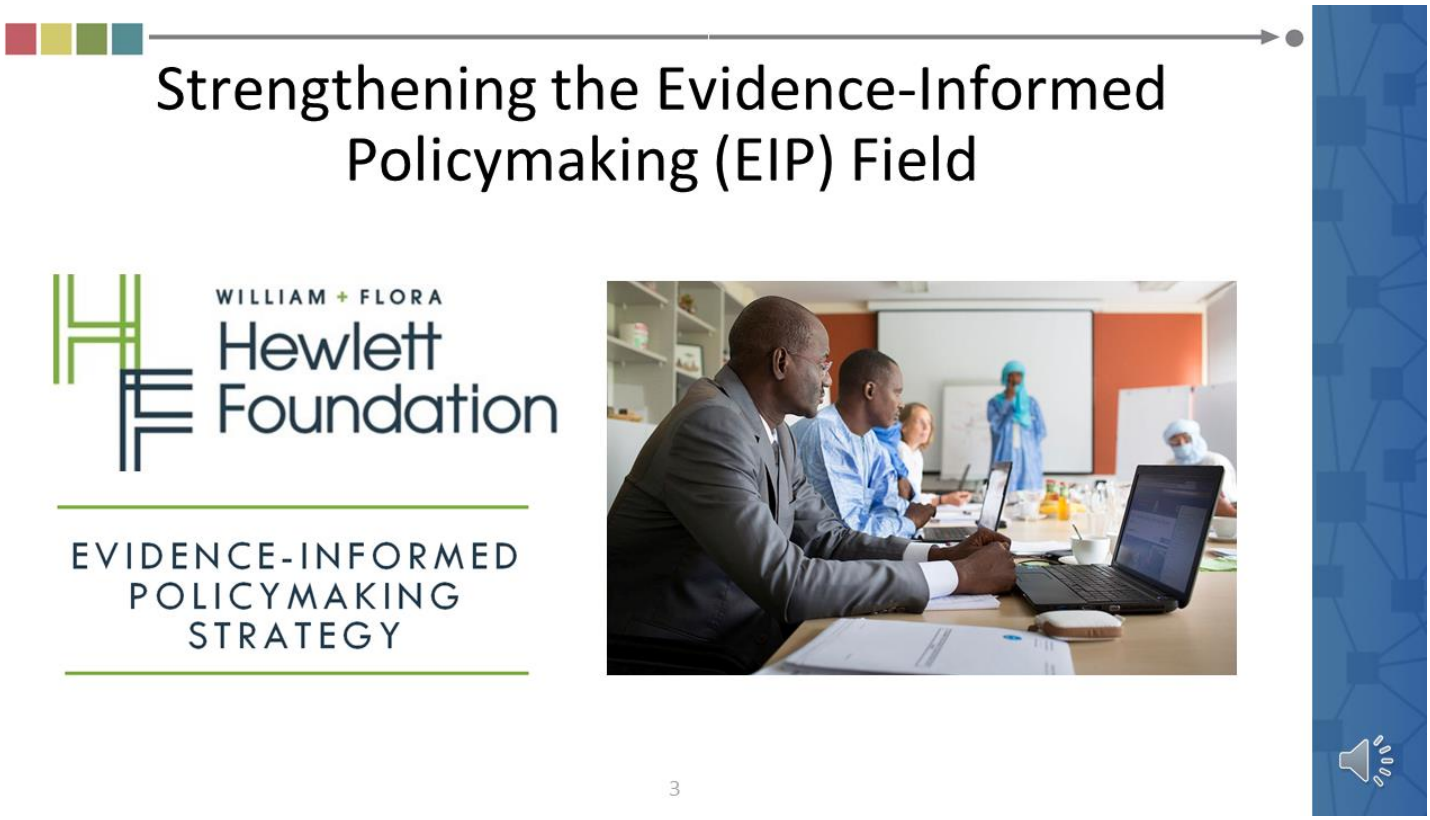
This document is a transcript of a video presentation by Bruce Hoppe and Sarah Lucas. All presentation materials are available at *http://connectiveassociates.com/2019/03/visualizing-eip-twitter/*.

[Bruce] Hello, I'm Bruce Hoppe with Connective Associates. My co-presenter Sarah Lucas and I welcome you to our brief presentation about using twitter mapping to help visualize and support a network of actors working for social change. We did this work with our colleague Claire Reinelt and the support of the William and Flora Hewlett Foundation.



[Bruce] This presentation might be useful to you if:

- You fund initiatives to connect, align, and activate networks around a shared vision
- You design communications strategies and want to understand who has influence in a communications ecosystem, such as Twitter
- You conduct social network analyses and want to learn how to find influencers and identify clusters of actors in a complex network ecosystem
- You are someone who works to increase government use of data and evidence for decision-making, and you are interested in getting a sense of other actors working toward similar aims and how they are connected.

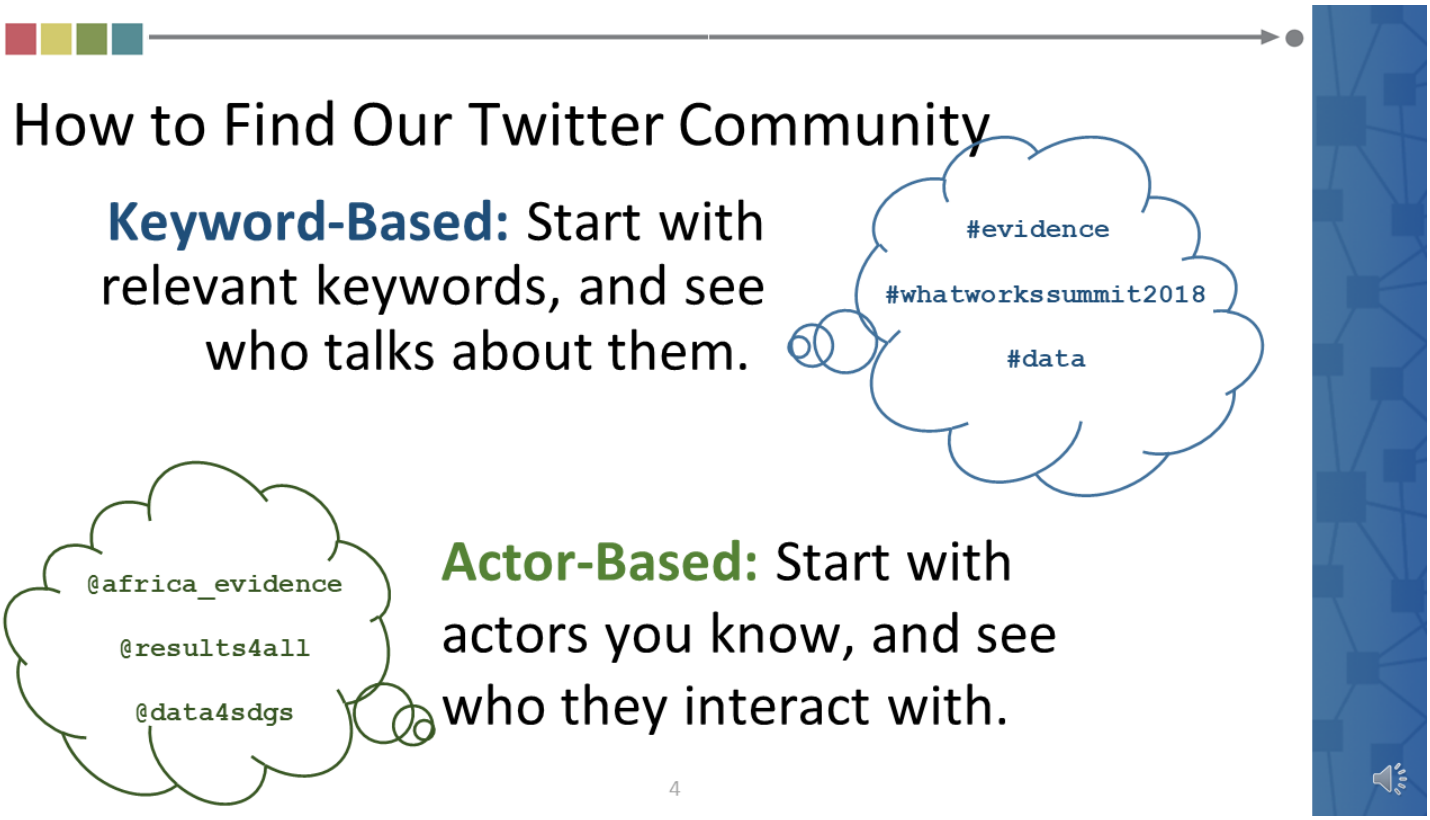


[Sarah] Hi, I'm Sarah Lucas with the William and Flora Hewlett Foundation. This work was commissioned by the Evidence-Informed Policymaking team at the William and Flora Hewlett Foundation.

Under our Evidence-Informed Policymaking strategy -- or EIP -- our goal is that governments use data and evidence to improve policymaking and routine decision-making. We were interested to learn what more we could do to strengthen and support the emerging EIP field -- that is the collection of organizations working to increase government use of evidence.

Before planning to strengthen the field, we wanted a better understanding of what this emerging field looks like. Who is in it, and how are they connected to each other.

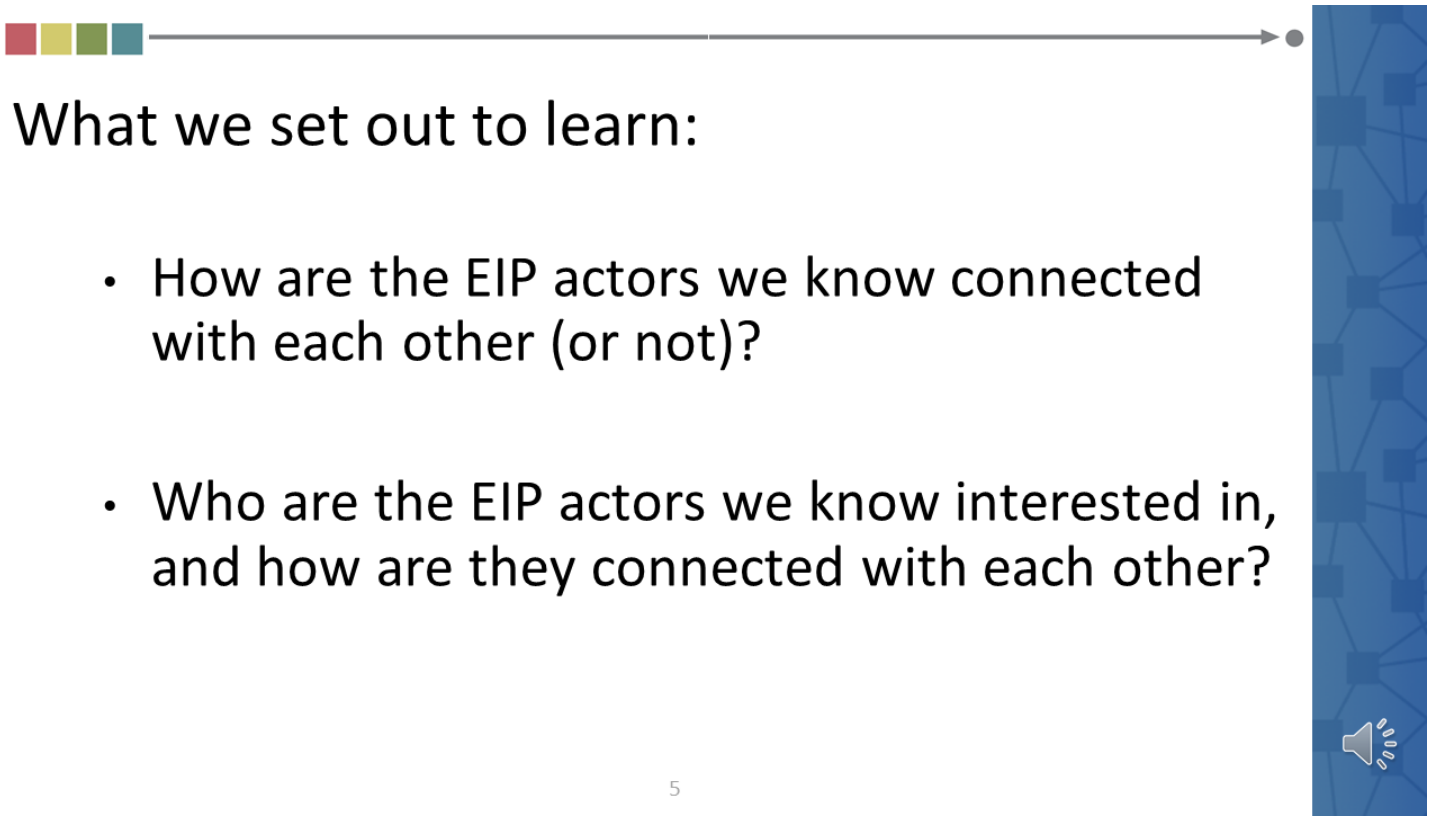
To do this, we turned to two experts in network mapping, Bruce Hoppe and Claire Reinelt. Together we looked at one of the tools available for network mapping -- Twitter.



[Bruce] We asked Sarah how to find EIP discussions on Twitter, and she provided both an initial list of users known to Hewlett and a short list of EIP-related keywords.

We hoped that keywords would enable us to see a more complete picture of who was tweeting about EIP without being limited by what Hewlett knew coming in, but we found our candidate keywords to be problematic. Some keywords like "evidence" were used so broadly that most of the results were not relevant to EIP. Other keywords like conference-based hashtags were used so narrowly that the results were entirely from users that Hewlett already knew.

We therefore decided to ignore keywords in our initial scan for the scope of the EIP twitter discussion, and to find the EIP community with an actor-based approach. Then after finding our EIP community we would analyze what community members talked about and arrive at keywords that way.



[Sarah] We set out to learn two main things from this Twitter mapping:

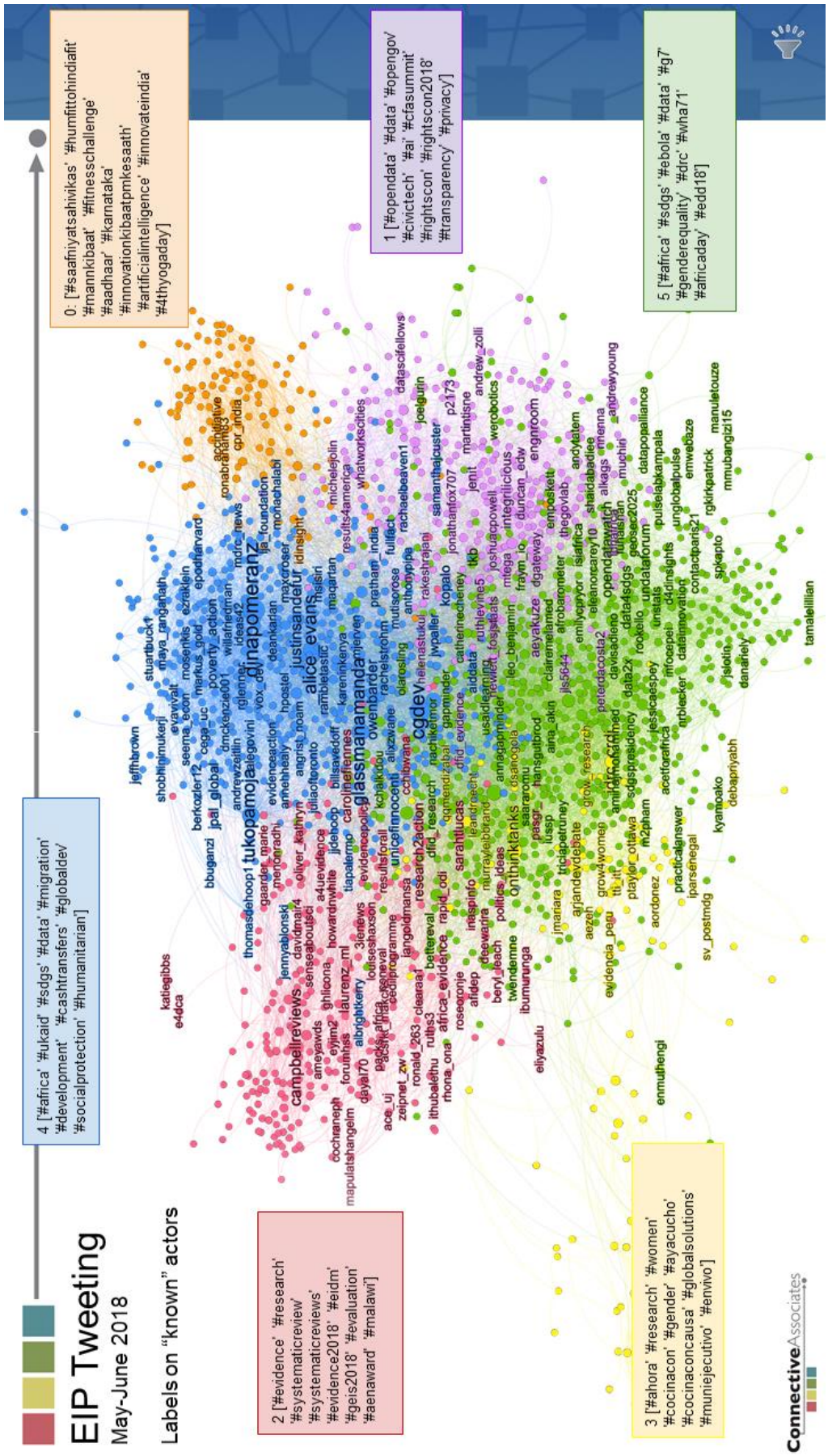
First, how are the EIP actors we know connected with each other (or not)? We started with a list of 250 EIP leaders that we know of. These are our grantees, scholars, government champions of evidence use, and other thought leaders.

We refer to this as the "known actor list." As you can see, an important limitation of this analysis is that it starts with people we already know! It's important to keep that in mind as you review the maps.

Second, who are these known actors are interested in, and how are they connected with each other? This is a larger pool of actors we refer to as the "enhanced known actor list." It is built by seeing who our known actors retweet and mention on twitter. Who are they paying attention to?

We know that a Twitter map -- especially one that starts just with the folks we already know -- cannot give us a complete or definitive picture of an emerging field of diverse actors, or anything about how they actually work. But what we were hoping for was some clues about who is working toward similar goals, and how they are connected with each other. With Bruce's help, we tried a variety of ways to get these clues.

We have two maps to show you. For each map, Bruce will explain what went into the making the slide and what you are seeking. Then I will share some thoughts on what the map might be suggesting in terms of the field.


EIP Tweeting
May-June 2018
Labels on "known" actors
0: ['#saafmiyatsahivikas' '#humfittohindiafit' '#mannkibaat' '#fitnesschallenge' '#aadhaar' '#karnataka' '#innovationkibaatpmkesaath' '#artificialintelligence' '#innovateindia' '#4thyogaday']
1 ['#opendata' '#data' '#opengov' '#civictech' '#ai' '#cfasummit' '#rightscon' '#rightscon2018' '#transparency' '#privacy']
2 ['#evidence' '#research' '#systematicreview' '#systematicreviews' '#evidence2018' '#eidm' '#geis2018' '#evaluation' '#aenaward' '#malawi']
3 ['#ahora' '#research' '#women' '#cocinacon' '#gender' '#ayacucho' '#cocinaconcausa' '#globalsolutions' '#muniejecutivo' '#envivo']
4 ['#africa' '#ukaid' '#sdgs' '#data' '#migration' '#development' '#cashtransfers' '#globaldev' '#socialprotection' '#humanitarian']
5 ['#africa' '#sdgs' '#ebola' '#data' '#g7' '#genderequality' '#drc' '#wha71' '#africaday' '#edd18']
ConnectiveAssociates

Supporting Evidence-Informed Policymaking: A Case Study in Twitter Network Visualization

[Bruce] We are now looking at our main evidence-informed policymaking network map. I'm going to explain the nuts and bolts of what is pictured here.

1. Each node, or dot, is a Twitter account. We started with about 250 accounts known to Hewlett. Then by tracking who those accounts are following, we developed an enhanced known actor list of about 2800 accounts. Most of those actively Tweet, and those are the nodes drawn here.

2. There are labels on some of the nodes. And those labels indicate the original 250 accounts known to Hewlett that we started with.

3. There are lines, or links, that connect some pairs of nodes. The lines are thin and light because there are so many of them (tens of thousands) and we don't want the map to look too overwhelming. Each link joins a pair of nodes that exchanged one or more Tweets during the three weeks we observed in May and June of 2018. In particular one account must have @-mentioned or retweeted the other account during that time, for a link to be mapped.

4. The nodes are drawn as a network map by a computer algorithm that puts nodes closer together when they are connected and farther apart when they are not connected. So when you see lots of nodes close together, that's because there are many links or tweets connecting them to each other. When nodes are far apart, say on opposite sides of the network map, then there are very few links connecting them.

5. Taking that idea just a bit further, we identify clusters. Each cluster is a group of nodes that are more connected to each other than to nodes outside the group. The map uses color to indicate different clusters. Note that clusters typically have inherently fuzzy boundaries. If you see a node that looks like it could belong to a couple different clusters, don't make too big a deal about how the clustering algorithm colored the node in one or the other.

6. You can see that some nodes are bigger or smaller, and the labels are sized to match the nodes. Node size and label size in this map indicate a metric called betweenness centrality. That is, roughly speaking, the extent to which a node is a bridge connecting the rest of the network. You can see the biggest nodes and labels in the map are in the blue cluster. These are twitter accounts that are well connected by @-mentions and retweets across multiple clusters.

7. Let's move from the network map to the colored boxes around the perimeter of this slide. Each box contains a list of hashtags that are important to a specific cluster.

8. To help understand what each cluster is about, we analyzed the content of every tweet for hashtags. We did this separately for each cluster. So, for example, looking at the purple cluster, we consider only tweets that are connecting purple nodes. What are the hashtags in those tweets? Of those hashtags, which ones were used by the most purple nodes? Those hashtags used most broadly by purple nodes are then listed in the purple box. And we do the same calculation for each cluster.

9. Finally, it's worth clarifying the relationship between clusters and hashtags. The clusters are based simply on how the nodes are connected to each other. Who is tweeting to whom, essentially. The cluster calculation does not look at hashtags. Only after we have calculated clusters do we then look at hashtags, and see, cluster by cluster, which hashtags get used the most in which clusters.

[Sarah] There is a lot to look at on this map, so we'll take it one section at a time.

1. Let's start with the red cluster on the upper left. Here you will see some names that are quite familiar to people working on impact evaluation and systematic reviews -- one form of evidence potentially useful for government policymaking. For example, you see the International Initiative for Impact Evaluation, 3ie, and big organizations producing systematic reviews like Campbell and Cochrane. You see a lot of connectivity here.
2. Now look down at lower right, the green section. This is largely a community working on data -- the Global Partnership for Sustainable Development Data (data4sdgs), UN Global Pulse, Data2X and the UN statistical community. This is a cluster of actors working on data for development and government decision-making. There are lots of connections within this community, but as you can see, if is very far away from the red cluster.
3. There is another cluster working on data in purple. What is the difference between the green and purple data clusters? Let's look to the hashtags. The purple cluster is mostly focused on open data and open government, whereas the green cluster is more focused on data for issues around Africa, the Sustainable Development goals, Gender. Two distinct communities. Open data and data for development. We know that these two communities exist, and they do overlap in some cases. You can see some folks right on the border between them, and some parts that remain quite distant.
4. Now look at big blue cluster in the middle. It appears to be a mix of think tanks, more impact evaluation organizations and some university centers. One thing to notice about this cluster is that some of the names in it are quite big. As Bruce explained, the bigger names suggest that these individuals are influencing or bridging across communities.

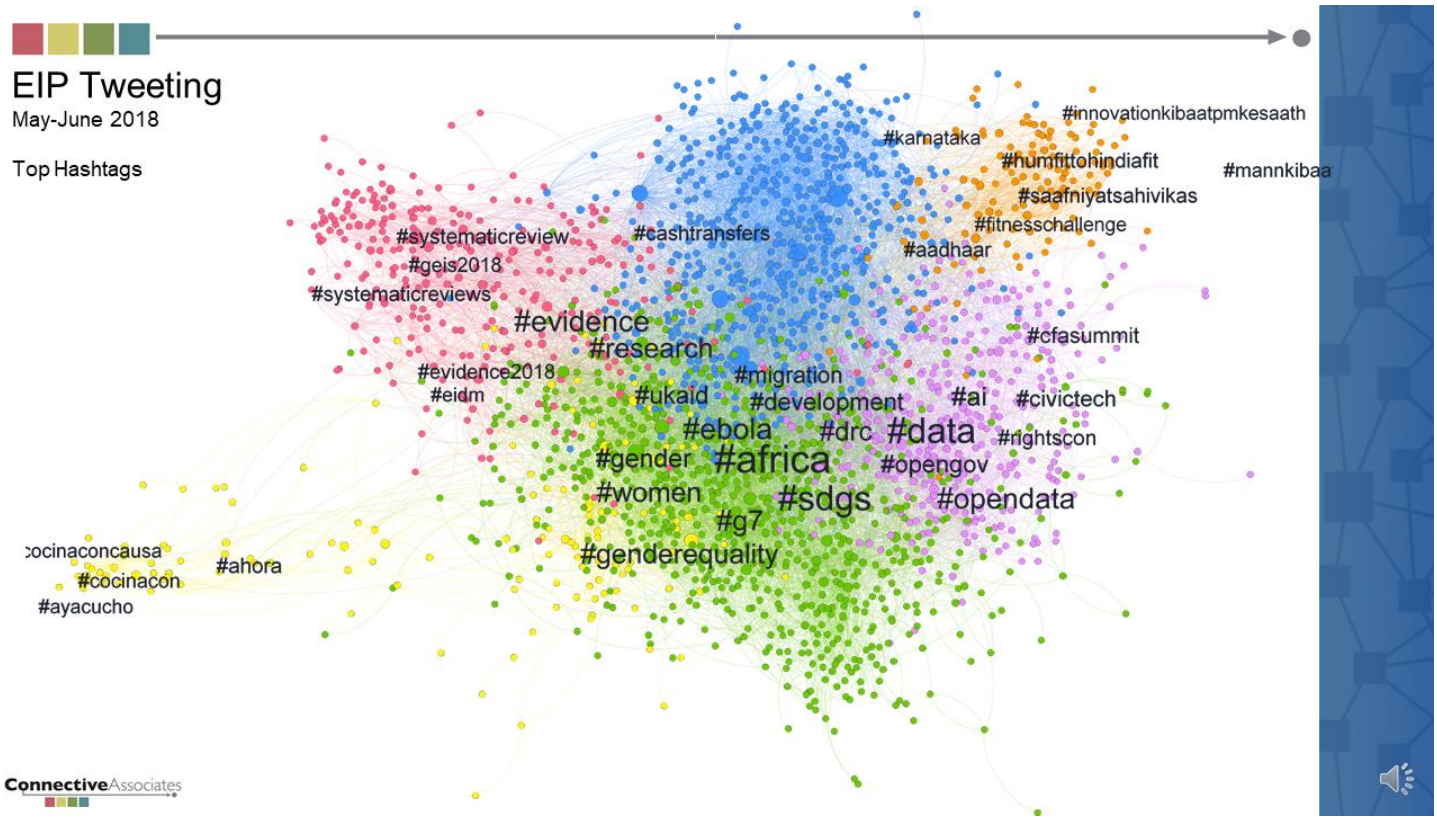
So, what are the main messages that emerge from this map that might help us to support the emerging EIP field?

First observation is that you have various tight-knit communities that are working with different kinds of data and evidence, towards a common goal of improving government decisions and policies, yet they are quite far away from each other. This raises questions about how we might bridge these communities that have similar EIP goals but use quite different methods.

The second observation is that we get some clues about who is starting to bridge across communities.

- One way is to see who the biggest names are. Those are the people and organizations that are most retweeted and @mentioned, so their voices are reaching across communities.
- Another way is to look at who is right at the border between the colored clusters. We see a few kinds of organizations on the borders.
    - We see some organizations that are have explicitly made EIP (rather than a specific type of evidence) central to their mandate, such as the Africa Evidence Network, African Institute for Development Policy, Results for All, INASP. (between red and green).
    - We see individuals and organizations that we know to be trying to deepen connections between the open data and data for development communities (green and purple) -- Claire Melamed from the GPSDD, Aidan Eyakuze from Twaweza, Shaida Baidee from Open Data Watch. Muchiri Nyaggah of LDRI and Al Kags of the Open Institute, both in Kenya.
    - We see some funders straddling different clusters -- USAID (green and blue), DFID (blue and green), UNICEF Innocenti (red, blue, green) IDRC (green and yellow), a number of us from Hewlett (Sarah red/green, Ruth green/purple/blue). This might suggest that funders can play a bridging role across communities.

So, this map gives us some interesting indications of who is in the EIP community, how they are clustered, and who might be best positioned to bridge across these clusters. Now let's turn to another way to understand this emerging field.



[Bruce] We're switching the map a bit now to focus on hashtags. I'm going to explain briefly how we constructed this map and then turn it back to Sarah.

The underlying nodes and links are exactly the same as the previous map. We have different labels now, indicating hashtags instead of account names. Recall that in the previous map, we looked at hashtags separately cluster by cluster. Here we instead take a unified approach and look at hashtags across the entire EIP Twitter network.

We start by taking the top 35 or so hashtags, measured by how many different accounts are using each hashtag. This is not *how often* the hashtag is used, but *how many different accounts* use the hashtag at least once. The hashtag labels are sized to indicate their popularity. Africa is the biggest hashtag, indicating that more accounts in this EIP network have used the Africa hashtag than any other hashtag.

The location of each hashtag indicates the approximate center of the network subgroup that is using that hashtag. This gives a rough sense of which nodes are using each hashtag. For example, hashtags Africa and Ebola are very close to each other, near the center of the network. Which accounts are using these hashtags? We can make an educated guess. We know that Africa is the most popular hashtag in the blue cluster, the most popular hashtag in the green cluster, *and* the most popular hashtag overall across the network. So we can infer that a large group of accounts spread across much of the EIP network (especially the blue and green clusters) are tweeting about Africa. It's a bit harder to say who is using the Ebola hashtag. We do know it's a very popular topic in the green cluster (where it's 3rd) but it is not in the top ten

hashtags in any other cluster. So we infer that a more compact group of accounts are tweeting about ebola, mostly in the green cluster.

Finally, when we see a hashtag on the periphery of the network, it's a bit easier to say who is using that hashtag. We can infer that the accounts using that hashtag are relatively tightly clustered around the location of the hashtag, with few accounts using the hashtag from other areas like the middle of the network, and probably no accounts using the hashtag from the opposite side.

[Sarah] This map gives us a sense of what people in these clusters are talking about, at least on Twitter.

One of the things that jumps out is to see who is talking primarily about types of evidence, versus those who are talking primarily about substantive topics.

For example, in the purple on the right, you see folks talking about data, open data, artificial intelligence (AI), civic tech. On the left in red, you see people talking about evidence, research, evaluation, systematic reviews. Whereas in the middle green you see folks talking more about issues -- women, sustainable development goals, Africa, gender, ebola, migration, etc. and presumably drawing on evidence to have these discussions.

This is obviously a huge generalization from this one map, but it does raise questions about how to build connections between communities that center their discussion around specific types of evidence, versus those that center their conversation around particular topics and champion evidence within that topic.

I also noticed in the blue section, which you'll recall had a lot of bridgers and influencers, didn't have a lot of hashtags. I asked Bruce about this and he thinks that is consistent with this section being populated by people connecting across communities. They are likely talking about the issues being discussed in all the other communities, so the hashtags show up predominantly in the other clusters, but presumably they all show up at a lower level in the blue.

So this map gives us some clues about why it might be that there is not more connection across these different clusters, and raises important questions about whether the best way to bridge across them is through substantive topic areas rather than types of evidence.

[Sarah] These two Twitter maps by no means offer a definitive analysis of the emerging Evidence-Informed Policymaking field. Alone, they don't tell us what we need to do to strengthen that field. But they have given us some clues about how communities define themselves, where connections are strongest and weakest, and who might be well-positioned to bridge across communities. Hopefully they also show how this kind of mapping can be useful to begin to understand a given network or field of actors, and generate some initial observations and questions for moving forward.

[Bruce] That concludes our presentation. Thank you for watching. The link here provides you with some additional resources. If you have any questions, we would love to hear from you. Thank you.

- Video presentation: http://connectiveassociates.com/2019/03/visualizing-eip-twitter/
- Guide to making & reading EIP network maps
- Twitter Directory of Initial List of EIP Actors

Contact us:

- Bruce Hoppe, Bruce@ConnectiveAssociates.com
- Sarah Lucas, SLucas@Hewlett.org, @sarahtlucas
- Claire Reinelt, Reinelt.Claire@gmail.com